PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : A61K 38/17, C07K 1/00, 14/00, C12P 21/00 — A1

(11) International Publication Number: WO 97/17988

(43) International Publication Date: 22 May 1997 (22.05.97)

(21) International Application Number: PCT/US96/18149

(22) International Filing Date: 13 November 1996 (13.11.96)

(30) Priority Data:
60/006,608 13 November 1995 (13.11.95) US
Not furnished 12 November 1996 (12.11.96) US

(71) Applicant: FIBROGEN, INC. [US/US]; 260 Littlefield Drive, South San Francisco, CA 94080 (US).

(72) Inventors: VUORIO, Eero; Hurtinkato 11C 18, FIN-20610 Turko (FI). MARTIN, George; 377 Everett Drive, Palo Alto, CA 94301 (US). NEFF, Thomas, B.; 190 Glenwood Avenue, Atherton, CA 94025 (US). ALA-KOKKO, Leena; University of Oulu, Kajaanintie 52A, FIN-90220 Oulu (FI).

(74) Agents: HALLUIN, Albert, P. et al.; Pennie & Edmonds, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CU, CZ, EE, FI, GE, HU, IL, IS, JP, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TR, TT, UA, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: TYPE IX COLLAGEN AND CHIMERAS

(57) Abstract

The present invention relates to novel collagens, polynucleotide sequences encoding these novel proteins, and to the use of these novel proteins and polynucleotides in the diagnosis and treatment of disease. The present invention further relates to specific collagens and derivatives, specifically fusion proteins of type IX collagen with type II and/or type XI collagens, and their use as therapeutic agents.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

TYPE IX COLLAGEN AND CHIMERAS

I. FIELD OF THE INVENTION

The present invention relates to novel human collagen proteins, polynucleotide sequences which encode these novel collagens proteins, and to the use of these novel proteins in the diagnosis and treatment of disease. More specifically, the present invention relates to polynucleotides encoding human α3(IX) collagen and derivatives hereof, fusion proteins of type IX collagen with type II and/or type XI collagen subunits and derivatives, and use of these proteins and polynucleotides as diagnostic and therapeutic agents.

II. BACKGROUND

Collagen fibrils, proteoglycan aggregates and glycoproteins are critical components of the cartilage extracellular matrix that, collectively, resist compression and the tensile and shear forces that are generated during articulation. Heinegird and Oldberg, *FASEB J. 3*:2042-2051 (1989); Mayne and Brewton, *Cartilage Degradation: Basic and Clinical Aspects* (Woessner, J.F. and Howell, D.S., eds.) Marcel Dekker, Inc., New York, pp. 81-108 (1993). Mutations in cartilage matrix genes that affect the biosynthesis, assembly or interactions between these various matrix components may contribute to degradation of the cartilage matrix and the loss of normal cartilage function. Mutations in human collagens have been shown to cause a series of chondrodysplasias ranging in severity from lethal achondrogenesis type II to Stickler arthroophthalmopathy and early onset familial osteoarthritis (reviewed by Spranger *et al.*, *Eur. J. Pediatr. 153*:56-65 (1994); Vikkula *et al.*, *Ann. Medicine 26*:107-114 (1994); Prockop and Kivirikko, *Annu. Rev. Biochem. 64*:403-434 (1995)).

Analyses of type IX collagen demonstrate this molecule is located on the surface of type II collagen-containing fibrils in hyaline cartilage and other tissues, including the vitreous humor (reviewed by Brewton and Mayne, *Extracellular Matrix Assembly and Structure* (Yurchenco, P.D., Birk, D.E., Mecham. R.P., eds) Academic Press, Inc., San Diego, pp. 129-170 (1994)). Type IX collagen is a heterotrimer composed of three polypeptide subunits: α1(IX), α2(IX) and α3(IX),

that are products of distinct genes and that contain alternating non-triple-helical or noncollagenous domains (NC1-4) and triple-helical or collagenous domains (COL1-3). The three polypeptide subunits are assembled into a mature collagen molecule with the structure α(IX)α2(IX)α3(IX) (van der Rest and Mayne, *Structure and Function of Collagen Types* (Mayne, R. and Burgeson, R., eds.) Academic Press, Orlando, FL, pp. 195-221 (1987). In addition to type II and type IX collagen, hyaline cartilage from a variety of sources also contains significant amounts of at least three other collagen molecules, types VI, X and XI. Thomas *et al.*, *Ann. Rheumat. Diseases 53*:488-496 (1994); Mayne and Brewton, *Cartilage Degradation: Basic and Clinical Aspects* (Woessner, J.F. and Howell, D.S., eds) Marcel Dekker, Inc., New York, pp. 81-108 (1993). Type XI collagen, like type IX collagen, is a heterotrimer composed of three different polypeptide subunits, α1(XI), α2(XI), and α3(XI). Collagen types XII and XIV were also isolated from bovine articular cartilage. Watt *et al.*, *J. Biol. Chem. 267*:20093-20099 (1992).

Native type IX collagen molecules interact with type II collagen molecules in a highly specific manner so that the domains NC1, COL1, NC2, COL2 and NC3 lie along the surface of the collagen fibril. The interactions between type IX and type II collagen are stabilized by multiple covalent crosslinks derived from specific lysine residues. *See* van der Rest and Mayne, *J. Biol. Chem. 263*:1615-1618 (1988); Shimokomaki *et al.*, *Ann. N.Y. Acad. Sci. 580*:1-7 (1990); Wu *et al.*, *J. Biol. Chem. 267*:23007-23014 (1992). The periodic localization of type IX collagen along type II collagen fibrils can be readily visualized by rotary shadowing because the collagenous domain COL3 and the large globular domain NC4 project from the surface of the fibril. Vaughan *et al.*, *J. Cell Biol. 106*:991-997 (1988); Shimokomaki *et al.*, *Ann. N.Y. Acad. Sci. 580*:1-7 (1990). In contrast, the type XI collagen heterotrimer is thought to reside in the central portion of the fibril. Mendler *et al.*, *J. Cell Biol. 108*:191-97 (1989).

Cloning and sequencing of the human type II collagen gene and the three human type XI collagen genes has been reported. The complete human type II collagen gene sequence was reported by Baldwin *et al.*, *Biochem. J. 262*:521-28 (1989) and by Su *et al.*, *Nucleic Acids Res. 17*:9473 (1989). Of the three type XI collagen subunits, the α3(XI) chain is believed to be the product of the type II collagen gene. Bernard *et al.*, *J. Biol. Chem. 263*:17159-66 (1988) discloses

cDNA sequence purportedly encoding the proα1(XI) collagen. Sequence coding for the α2(XI) gene was reported by Kimura *et al.*, *J. Biol. Chem. 264*:13910-16 (1989).

The genes encoding the three chains of type IX collagen are excellent candidates for chondrodysplasias and degenerative disorders that affect the joints and/or vitreous humor because type IX collagen is a significant structural molecule in both of these tissues. Therefore, cloning of the genes encoding the three type IX collagen subunits has been the object of intensive research. Muragaki *et al.*, *Eur. J. Biochem. 192*:703-8 (1990), presented the complete cDNA sequence of both alternative transcripts from the human α1(IX) gene. The majority of the human α2(IX) collagen cDNA was reported by Perala *et al.*, *FEBS Lett. 319*:177-80 (1993) and completed by Warman *et al.*, *Genomics 23*:158-62 (1994). The complete human sequence for the α3(IX) subunit was unavailable until recently. As described in the concurrently filed application (U.S. Application to be assigned), Drs. R.W. Brewton and R. Mayne have identified and characterized the full length sequence corresponding to human α3(IX). The information contained in the Brewton and Mayne provisional application is incorporated herein by reference.

Experiments utilizing transgenic mice suggest that type IX collagen plays an important role in maintaining the integrity of hyaline cartilage. Animals that either express a minigene carrying a deletion in the α1(IX) chain (Nakata *et al.*, *Proc. Natl. Acad. Sci. U.S.A. 90*:2870-2874 (1993)) or that carry disrupted α1(IX) genes (Fèssler *et al.*, *Proc. Natl. Acad. Sci. U.S.A. 91*:5070-5074 (1994)) develop degenerative joint disease that resembles human osteoarthritis. The importance of type IX collagen in human disease was verified by the identification of a mutation in *COL9A2* (Muragaki *et al.*, submitted for publication, (1995)) that results in the skipping of exon 3 and that causes Multiple Epiphyseal Dysplasia (EDM2).

III. SUMMARY OF THE INVENTION

The present invention relates to novel collagen derivative proteins and the polynucleotide sequences which encode them. Also described herein are methods for diagnosis of diseases resulting from abnormalities in collagen synthesis or structure.

One aspect of this invention is the discovery that fusion proteins of human type IX collagen may be produced in which a human type IX collagen subunit is covalently linked to human type II collagen and/or a human type XI collagen subunit. In one embodiment of this invention, the fusion proteins are recombinantly produced as chimeras by linking polynucleotide coding sequence for a human type IX collagen subunit in frame to polynucleotide coding sequence for human type II collagen and/or a human type XI collagen. This chimeric coding sequence is inserted into an expression vector and used to transform appropriate host cells. The host cells are then induced to express the chimeric coding sequence and thereby produce the chimeric collagen fusion proteins. These fusion proteins are useful in the treatment of collagen related diseases and conditions.

The present invention also relates, in part, to nucleotide sequences and expression vectors encoding the chimeric collagens of the invention.

Also disclosed herein are methods of treatment of diseases or conditions associated with abnormalities in collagen production or autoimmunity to collagen. Such abnormalities can result in, for example, rheumatoid arthritis, osteoarthritis, reactive arthritis, autoimmune hearing disease, cartilage inflammation due to bacterial or viral infections (e.g. Lyme's disease), parasitic disease, bursitis, corneal diseases, and ankylosing spondylitis (fusion of the spine). The novel proteins of the invention are used in these methods of treating collagen related diseases.

IV. BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Salt fractionation of recombinant human type IX collagen heterotrimer of α1(IX), α2(IX), and α3(IX) in nonreducing 8% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

Figure 2. Salt fractionation of recombinant human type IX collagen heterotrimer of α1(IX), α2(IX), and α3(IX) in reducing 10% SDS-PAGE.

V. DETAILED DESCRIPTION

The present invention relates to the polynucleotide and nucleic acid sequences encoding the recombinant fusion proteins of type IX collagen derivatives with type II collagen protein and/or type XI collagen protein and such fusion

proteins. Also within the scope of the invention are methods of using these collagen fusion proteins to treat collagen related disorders and conditions.

A. **Definitions**

The term "collagen subunit" refers to the amino acid sequence of one subunit of a collagen protein encoded by a single gene, as well as derivatives, including deletion derivatives, conservative substitutions, etc.

A "fusion protein" is a protein in which peptide sequences from different proteins are covalently linked together.

The term "chimera" or "chimeric" refers to fusion protein produced by operably linking polynucleotide coding sequence for two or more collagen protein subunits in frame and recombinantly expressing the linked coding sequence as a single peptide chain.

"Active human type IX collagen" refers to the native trimeric protein complex, and may be recombinantly produced.

The phrase "stringent conditions" as used herein refers to those hybridizing conditions that (1) employ low ionic strength and high temperature for washing, for example, 0.015 M NaCl/0.0015 M sodium citrate/0.1% SDS at 50°C; (2) employ during hybridization a denaturing agent such as formamide, for example, 50% (vol/vol) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM NaCl, 75 mM sodium citrate at 42°C; or (3) employ 50% formamide, 5 x SSC (0.75 M NaCl, 0.075 M Sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 g/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 x SSC and 0.1% SDS.

In accordance with the invention, any nucleotide sequence which encodes the amino acid sequence of claimed fusion protein can be used to generate recombinant molecules which direct the expression of the fusion protein.

The term "purified" as used herein in reference to collagens denotes that the indicated molecules are present in the substantial absence of other biological macromolecules, *e.g.*, polynucleotides, proteins, and the like. The term "purified" as used herein preferably means at least 95% by weight, more preferably at least 99.8% by weight, of the indicated biological macromolecules present (but water,

buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons can be present). The term "isolated" as used herein refers to a protein molecule separated not only from other proteins that are present in the natural source of the protein, but also from other proteins, and preferably refers to a protein found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass proteins present in their natural source.

B. Expression Of The Collagen Fusion Proteins Of The Invention

1. Coding Sequences

In accordance with the invention, polynucleotide sequences which encode type IX, type II, and type XI collagen proteins, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of fusion proteins of type IX collagen subunit with type II collagen and/or type XI collagen subunits, or a functional equivalent thereof, in appropriate host cells. Such collagen polynucleotide sequences, as well as other polynucleotides which selectively hybridize to at least a part of such collagen polynucleotides or their complements, may also be used in nucleic acid hybridization assays, Southern and Northern blot analyses, etc.

Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence, may be used in the practice of the invention for the cloning and expression of these collagen proteins. Such DNA sequences include those which are capable of hybridizing to the appropriate human collagen sequence under stringent conditions.

Altered DNA sequences which may be used in accordance with the invention include deletions, additions or substitutions of different nucleotide residues resulting in a sequence that encodes the same or a functionally equivalent gene product. The gene product itself may contain deletions, additions or substitutions of amino acid residues within a collagen sequence, which result in a silent change thus producing a functionally equivalent collagen. Such amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility,

hydrophobicity, hydrophilicity, and/or the amphipatic nature of the residues involved. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; amino acids with uncharged polar head groups having similar hydrophilicity values include the following: leucine, isoleucine, valine; glycine, alanine; asparagine, glutamine; serine, threonine; phenylalanine, tyrosine.

The DNA sequences of the invention may be engineered in order to alter the collagen coding sequence for a variety of ends including but not limited to alterations which modify processing and expression of the gene product. For example, alternative secretory signals may be substituted for the native human secretory signal and/or mutations may be introduced using techniques which are well known in the art, *e.g.*, site-directed mutagenesis, to insert new restriction sites, to alter glycosylation patterns, phosphorylation, etc. Additionally, when expressing in non-human cells, the polynucleotides encoding the collagens of the invention may be modified in the silent position of any triplet amino acid codon so as to better conform to the codon preference of the particular host organism.

In another embodiment of the invention, a collagen sequence may be ligated to a heterologous sequence to encode a fusion protein. For example, a fusion protein may be engineered to contain a cleavage site located between an α3(IX) collagen sequence and the heterologous protein sequence, so that the α3(IX) collagen may be cleaved away from the heterologous moiety.

In a particularly preferred embodiment, chimeric fusion proteins are constructed by ligating sequence encoding a type IX collagen subunit, or derivative thereof, to sequence encoding type II collagen and/or type XI collagen subunit. One of skill in the art will recognize that several techniques are available which allow one to ligate all or any part of the coding sequence of a type IX collagen subunit to all or any part of the coding sequences for type II and type XI collagens. For example, one may ligate coding sequences together at appropriately chosen restriction endonuclease sites. However, in order to ensure that the coding sequences of the chosen collagens are ligated in the correct translational frame, it may be necessary to engineer restriction sites by site specific mutagenesis. An even more advantageous method to join two or more polynucleotide sequences makes use of the polymerase chain reaction and appropriately designed primers as

described in Section 3.17.1 of Ausubel *et al.*, *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley Interscience, N.Y., (1990). Using this method, one of skill in the art may join two or more polynucleotide sequences in any configuration.

In an alternate embodiment of the invention, the coding sequence of the collagens of the invention could be synthesized in whole or in part, using chemical methods well known in the art. See, for example, Caruthers *et al.*, *Nuc. Acids Res. Symp. Ser.* 7:215-233 (1980); Crea and Horn, *Nuc. Acids Res. 9(10)*:2331 (1980); Matteucci and Caruthers, *Tetrahedron Letters 21*:719 (1980); and Chow and Kempe, *Nuc. Acids Res. 9(12)*:2807-2817 (1981). Alternatively, the protein itself could be produced using chemical methods to synthesize the desired collagen amino acid sequence at least in part. For example, peptides can be synthesized by solid phase techniques, cleaved from the resin, and purified by preparative high performance liquid chromatography. (e.g., see Creighton, *Proteins Structures And Molecular Principles*, W.H. Freeman and Co., N.Y., pp. 50-60 (1983). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (*e.g.*, the Edman degradation procedure; see Creighton, *Proteins, Structures and Molecular Principles*, W.H. Freeman and Co., N.Y., pp. 34-49 (1983).

In order to express the collagens of the invention, the nucleotide sequence encoding the collagen, or a functional equivalent, is inserted into an appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

2. Expression Systems

Methods which are well known to those skilled in the art can be used to construct expression vectors containing a collagen coding sequence for the collagens of the invention and appropriate transcriptional/translational control signals. These methods include *in vitro* recombinant DNA techniques, synthetic techniques and *in vivo* recombination/genetic recombination. See, for example, the techniques described in Maniatis *et al.*, *Molecular Cloning A Laboratory Manual*, Cold Spring Harbor Laboratory, N.Y. (1989) and Ausubel *et al.*, *Current*

Protocols in Molecular Biology, Greene Publishing Associates and Wiley Interscience, N.Y. (1989).

A variety of host-expression vector systems may be utilized to express a collagen coding sequence. These include but are not limited to microorganisms such as bacteria transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing a collagen coding sequence; yeast transformed with recombinant yeast expression vectors containing a collagen coding sequence; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing sequence encoding the collagens of the invention; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (*e.g.*, Ti plasmid) containing a collagen coding sequence; or animal cell systems. Additionally, the collagens of the invention may be expressed in transgenic non-human animals wherein the desired collagen product may be recovered from the milk of the transgenic animal. The expression elements of these systems vary in their strength and specificities. Depending on the host/vector system utilized, any of a number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used in the expression vector. For example, when cloning in bacterial systems, inducible promoters such as pL of bacteriophage l, plac, ptrp, ptac (ptrp-lac hybrid promoter) and the like may be used; when cloning in insect cell systems, promoters such as the baculovirus polyhedron promoter may be used; when cloning in plant cell systems, promoters derived from the genome of plant cells (*e.g.*, heat shock promoters; the promoter for the small subunit of RUBISCO; the promoter for the chlorophyll a/b binding protein) or from plant viruses (*e.g.*, the 35S RNA promoter of CaMV; the coat protein promoter of TMV) may be used; when cloning in mammalian cell systems, promoters derived from the genome of mammalian cells (*e.g.*, metallothionein promoter) or from mammalian viruses (*e.g.*, the adenovirus late promoter; the vaccinia virus 7.5 K promoter) may be used; when generating cell lines that contain multiple copies of a collagen DNA, SV40-, BPV- and EBV-based vectors may be used with an appropriate selectable marker.

In bacterial systems a number of expression vectors may be advantageously selected depending upon the use intended for the collagen expressed. For example, when large quantities of the collagens of the invention are to be produced for the generation of antibodies, vectors which direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include but are not limited to the *E. coli* expression vector pUR278 (Ruther *et al.*, *EMBO J.* 2:1791 (1983)), in which the collagen coding sequence may be ligated into the vector in frame with the lac Z coding region so that a hybrid AS-lac Z protein is produced; pIN vectors (Inouye & Inouye, *Nucleic Acids Res.* 13:3101-3109 (1985); Van Heeke & Schuster, *J. Biol. Chem.* 264:5503-5509 (1989)); and the like. pGEX vectors may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety.

A preferred expression system is a yeast expression system. In yeast, a number of vectors containing constitutive or inducible promoters may be used. For a review see, *Current Protocols in Molecular Biology, Vol. 2*, Ed. Ausubel *et al.*, Greene Publish. Assoc. & Wiley Interscience, Ch. 13 (1988); Grant *et al.*, *Expression and Secretion Vectors for Yeast, in Methods in Enzymology*, Ed. Wu & Grossman, Acad. Press, N.Y. *153*:516-544 (1987); Glover, *DNA Cloning, Vol. II*, IRL Press, Wash., D.C., Ch. 3 (1986); and Bitter, Heterologous Gene Expression in Yeast, *Methods in Enzymology*, Eds. Berger & Kimmel, Acad. Press, N.Y. *152*:673-684 (1987); and *The Molecular Biology of the Yeast Saccharomyces*, Eds. Strathern *et al.*, Cold Spring Harbor Press, Vols. I and II (1982).

A particularly preferred system useful for cloning and expression of the collagen proteins of the invention uses host cells from the yeast *Pichia*. Species of non-*Saccharomyces* yeast such as *Pichia pastoris* appear to have special advantages in producing high yields of recombinant protein in scaled up procedures. Additionally, a *Pichia* expression kit is available from Invitrogen Corporation (San Diego, CA).

There are a number of methanol responsive genes in methylotrophic yeasts such as *Pichia pastoris*, the expression of each being controlled by methanol responsive regulatory regions (also referred to as promoters). Any of such methanol responsive promoters are suitable for use in the practice of the present invention. Examples of specific regulatory regions include the promoter for the primary alcohol oxidase gene from *Pichia pastoris* AOX1, the promoter for the secondary alcohol oxidase gene from *P. pastoris* AXO2, the promoter for the dihydroxyacetone synthase gene from *P. pastoris* (DAS), the promoter for the P40 gene from *P. pastoris*, the promoter for the catalase gene from *P. pastoris*, and the like.

Typical expression in *Pichia pastoris* is obtained by the promoter from the tightly regulated AOX1 gene. See Ellis *et al., Mol. Cell. Biol. 5*:1111 (1985) and U.S. Patent No. 4,855,231. This promoter can be induced to produce high levels of recombinant protein after addition of methanol to the culture. By subsequent manipulations of the same cells, expression of genes for the collagens of the invention described herein is achieved under conditions where the recombinant protein is adequately hydroxylated by prolyl 4-hydroxylase and, therefore, can fold into a stable helix that is required for the normal biological function of the proteins in forming fibrils.

Another particularly preferred yeast expression system makes use of the methylotrophic yeast *Hansenula polymorpha*. Growth on methanol results in the induction of key enzymes of the methanol metabolism, namely MOX (methanol oxidase), DAS (dihydroxyacetone synthase) and FMHD (formate dehydrogenase). These enzymes can constitute up to 30-40% of the total cell protein. The genes encoding MOX, DAS, and FMDH production are controlled by very strong promoters which are induced by growth on methanol and repressed by growth on glucose. Any or all three of these promoters may be used to obtain high level expression of heterologous genes in *H. polymorpha*. The gene encoding a collagen of the invention is cloned into an expression vector under the control of an inducible *H. polymorpha* promoter. If secretion of the product is desired, a polynucleotides encoding a signal sequence for secretion in yeast, such as the *S. cerevisiae* prepro-mating factor α1, is fused in frame with the coding sequence for the collagen of the invention. The expression vector preferably contains an

auxotrophic marker gene, such as URA3 or LEU2, which may be used to complement the deficiency of an auxotrophic host.

The expression vector is then used to transform *H. polymorpha* host cells using techniques known to those of skill in the art. An interesting and useful feature of *H. polymorpha* transformation is the spontaneous integration of up to 100 copies of the expression vector into the genome. In most cases, the integrated DNA forms multimers exhibiting a head-to- tail arrangement. The integrated foreign DNA has been shown to be mitotically stable in several recombinant strains, even under non-selective conditions. This phenomena of high copy integration further adds to the high productivity potential of the system.

In cases where plant expression vectors are used, the expression of sequences encoding the collagens of the invention may be driven by any of a number of promoters. For example, viral promoters such as the 35S RNA and 19S RNA promoters of CaMV (Brisson *et al.*, *Nature 310*:511-514 (1984), or the coat protein promoter of TMV (Takamatsu *et al.*, *EMBO J. 6*:307-311 (1987)) may be used; alternatively, plant promoters such as the small subunit of RUBISCO (Coruzzi *et al.*, *EMBO J. 3*:1671-1680 (1984); Broglie *et al.*, *Science 224*:838-843 (1984); or heat shock promoters, *e.g.*, soybean hsp17.5-E or hsp17.3-B (Gurley *et al.*, *Mol. Cell. Biol. 6*:559-565 (1986) may be used. These constructs can be introduced into plant cells using Ti plasmids, Ri plasmids, plant virus vectors, direct DNA transformation, microinjection, electroporation, etc. For reviews of such techniques see, for example, Weissbach & Weissbach, *Methods for Plant Molecular Biology, Academic Press*, NY, Section VIII, pp. 421-463 (1988); and Grierson & Corey, *Plant Molecular Biology*, 2d Ed., Blackie, London, Ch. 7-9 (1988).

An alternative expression system which could be used to express the collagens of the invention is an insect system. In one such system, *Autographa californica* nuclear polyhidrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. Coding sequence for the collagens of the invention may be cloned into non- essential regions (for example the polyhedron gene) of the virus and placed under control of an AcNPV promoter (for example, the polyhedron promoter). Successful insertion of a collagen coding sequence will result in inactivation of the polyhedron gene and

production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedron gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed. (*e.g.*, see Smith *et al.*, *J. Virol. 46*:584 (1983); Smith, U.S. Patent No. 4,215,051).

In mammalian host cells, a number of viral based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, coding sequence for the collagens of the invention may be ligated to an adenovirus transcription/translation control complex, *e.g.*, the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non- essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing collagen in infected hosts. (*e.g.*, See Logan & Shenk, *Proc. Natl. Acad. Sci. (USA) 81*:3655-3659 (1984)). Alternatively, the vaccinia 7.5 K promoter may be used. (*See, e.g.*, Mackett *et al.*, *Proc. Natl. Acad. Sci. (USA) 79*:7415- 7419 (1982); Mackett *et al.*, *J. Virol. 49*:857-864 (1984); Panicali *et al.*, *Proc. Natl. Acad. Sci. 79*:4927-4931 (1982).

Specific initiation signals may also be required for efficient translation of inserted collagen coding sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where the entire collagen gene, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of a collagen coding sequence is inserted, exogenous translational control signals, including the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the collagen coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (see Bittner *et al.*, *Methods in Enzymol. 153*:516-544 (1987)).

One preferred expression system for the recombinant production of the collagens of the invention is in transgenic non-human animals, wherein the desired

collagen may be recovered from the milk of the transgenic animal. Such a system is constructed by operably linking the DNA sequence encoding the collagens of the invention to a promoter and other required or optional regulatory sequences capable of effecting expression in mammary glands. Likewise, required or optional post-translational enzymes may be produced simultaneously in the target cells, employing suitable expression systems, as disclosed in, *inter alia*, U.S. Application, Serial No. 08/037,728, operable in the targeted milk protein producing mammary gland cells.

For expression in milk, the promoter of choice would preferably be from one of the abundant milk- specific proteins, such as alpha S1-casein, or b-lactoglobulin. For example, 5' and 3' regulatory sequences of alpha S1-casein have been successfully used for the expression of the human lactoferrin cDNA, and similarly, the b-lactoglobin promoter has effected the expression of human antitrypsin gene fragments in sheep milk producing cells. Wright *et al.*, *Biotechnology* 9:830-833 (1991). In transgenic goats, the whey acid promoter has been used for the expression of human tissue plasminogen activator, resulting in the secretion of human tissue plasminogen activator in the milk of the transgenics. Ebert *et al.*, *Biotechnology* 9:835-838 (1991). Using such expression systems, animals are obtained which secrete the collagens of the invention into milk. Using procedures well-known by those of the ordinary skill in the art, the gene encoding the desired collagen chain can simply be ligated to suitable control sequences which function in the mammary cells of the chosen animal species. Expression systems for the genes encoding the required post-translational enzymes are constructed analogously.

Preferably, the collagens of the invention are expressed as secreted proteins. When the engineered cells used for expression of the proteins are non- human host cells, it is often advantageous to replace the human secretory signal peptide of the collagen protein with an alternative secretory signal peptide which is more efficiently recognized by the host cell's secretory targeting machinery. The appropriate secretory signal sequence is particularly important in obtaining optimal fungal expression of mammalian genes. For example, in methylotrophic yeasts, a DNA sequence encoding the in-reading frame *S. cerevisiae* a- mating factor pre-pro sequence may be inserted at the amino-terminal end of the coding sequence. The

aMF pre-pro sequence is a leader sequence contained in the aMF precursor molecule, and includes the lys-arg encoding sequence which is necessary for proteolytic processing and secretion (*see, e.g.*, Brake *et al.*, *Proc. Nat'l. Acad. Sci. USA, 81*:4642 (1984)).

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (*e.g.*, glycosylation) and processing (*e.g.*, cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post- translational processing and modification of proteins. Appropriate cells lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, 293, WI38, etc. Additionally, host cells may be engineered to express various enzymes to ensure the proper processing of the collagen molecules. For example, the gene for prolyl-4-hydroxylase may be coexpressed with the collagen gene in the host cell.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express the collagens of the invention may be engineered. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with collagen encoding DNA controlled by appropriate expression control elements (*e.g.*, promoter, enhancer, sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines which express a desired collagen.

A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler *et al.*, *Cell 11*:223 (1977)), hypoxanthine-guanine phosphoribosyltransferase (Szybalska & Szybalski, *Proc. Natl. Acad. Sci. USA 48*:2026 (1962)), and adenine phosphoribosyltransferase (Lowy *et al.*, *Cell 22*:817 (1980)) genes can be employed in tk-, hgprt- or aprt- cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for dhfr, which confers resistance to methotrexate (Wigler *et al.*, *Natl. Acad. Sci. USA 77*:3567 (1980); O'Hare *et al.*, *Proc. Natl. Acad. Sci. USA 78*:1527 (1981)); gpt, which confers resistance to mycophenolic acid (Mulligan & Berg, *Proc. Natl. Acad. Sci. USA 78*:2072 (1981)); neo, which confers resistance to the aminoglycoside G-418 (Colberre- Garapin *et al.*, *J. Mol. Biol. 150*:1 (1981)); and hygro, which confers resistance to hygromycin (Santerre *et al.*, *Gene 30*:147 (1984)). Recently, additional selectable genes have been described, namely trpB, which allows cells to utilize indole in place of tryptophan; hisD, which allows cells to utilize histinol in place of histidine (Hartman & Mulligan, *Proc. Natl. Acad. Sci. USA 85*:8047 (1988)); and ODC (ornithine decarboxylase) which confers resistance to the ornithine decarboxylase inhibitor, 2-(difluoromethyl)-DL-ornithine, DFMO (McConlogue L., *In: Current Communications in Molecular Biology*, Cold Spring Harbor Laboratory, Ed.) (1987).

C. Identification of Transfectants or Transformants that Express the Collagen Proteins of the Invention and Purification of the Expressed Proteins

The host cells which contain the coding sequence and which express the biologically active gene product may be identified by at least four general approaches; (a) DNA-DNA or DNA-RNA hybridization; (b) the presence or absence of "marker" gene functions; (c) assessing the level of transcription as measured by the expression of collagen mRNA transcripts in the host cell; and (d) detection of the gene product as measured by immunoassay or by its biological activity.

In the first approach, the presence of the collagen coding sequence inserted in the expression vector can be detected by DNA-DNA or DNA-RNA

hybridization using probes comprising nucleotide sequences that are homologous to the collagen coding sequence, respectively, or portions or derivatives thereof.

In the second approach, the recombinant expression vector/host system can be identified and selected based upon the presence or absence of certain "marker" gene functions (*e.g.*, thymidine kinase activity, resistance to antibiotics, resistance to methotrexate, transformation phenotype, occlusion body formation in baculovirus, etc.). For example, if the collagen coding sequence is inserted within a marker gene sequence of the vector, recombinant cells containing collagen coding sequence can be identified by the absence of the marker gene function. Alternatively, a marker gene can be placed in tandem with the collagen sequence under the control of the same or different promoter used to control the expression of the collagen coding sequence. Expression of the marker in response to induction or selection indicates expression of the collagen coding sequence.

In the third approach, transcriptional activity of the collagen coding region can be assessed by hybridization assays. For example, RNA can be isolated and analyzed by Northern blot using a probe homologous to the collagen coding sequence or particular portions thereof. Alternatively, total nucleic acids of the host cell may be extracted and assayed for hybridization to such probes.

In the fourth approach, the expression of a collagen protein product can be assessed immunologically, for example by Western blots, immunoassays such as radioimmuno-precipitation, enzyme-linked immunoassays and the like.

The expressed collagen of the invention, which is preferably secreted into the culture medium, is purified to homogeneity, e.g., by chromatography. In one embodiment, the recombinant collagen protein is purified by size exclusion chromatography. However, other purification techniques known in the art can also be used, including ion exchange chromatography, and reverse-phase chromatography.

D. Uses of The Collagens of the Invention and Engineered Cell Lines

1. Antibody Production and Screening

Various procedures known in the art may be used for the production of antibodies to epitopes of the recombinantly produced collagens.

Such antibodies include but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by a Fab expression library.

For the production of antibodies, various host animals may be immunized by injection with a collagen protein including but not limited to rabbits, mice, rats, etc. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum*.

Monoclonal antibodies to a collagen may be prepared by using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include but are not limited to the hybridoma technique originally described by Koehler and Milstein, (*Nature*, *256*:495-497 (1975)), the human B-cell hybridoma technique (Kosbor *et al.*, *Immunology Today*, *4*:72 (1983)); Cote *et al.*, *Proc. Natl. Acad. Sci.*, *80*:2026-2030 (1983) and the EBV-hybridoma technique (Cole *et al.*, *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96 (1985). In addition, techniques developed for the production of "chimeric antibodies" (Morrison *et al.*, *Proc. Natl. Acad. Sci.* *81*:6851-6855 (1984); Neuberger *et al.*, *Nature*, *312*:604-608 (1984); Takeda *et al.*, *Nature 314*:452-454 (1985)) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. Alternatively, techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce a collagen- specific single chain antibodies.

Antibody fragments which contain deletions of specific binding sites may be generated by known techniques. For example, such fragments include but are not limited to: the F(ab')2 fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed (Huse *et al.*, *Science 246*:1275-1281 (1989)) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity for the collagen of interest.

2. Therapeutic Uses Of The Collagen Proteins Of The Invention

Another aspect of the invention is to provide methods of treating immune system-mediated diseases using the collagen proteins of the invention. The terms "treatment" or "treating" as used herein with reference to a disease refer both to prophylaxis and to the amelioration of symptoms already present in an individual. It will be appreciated by the person of ordinary skill in the art that a treatment need not be completely effective in preventing the onset of a disease or in reducing the symptoms associated with the disease. Any reduction of the severity of symptoms, delay in the onset of symptoms, or delay in the progression of severity of symptoms is desirable to a patient. Persons at risk of developing a given immune system-mediated disease may be treated prophylactically based on any of a variety of factors suggesting the possible onset of an immune system- mediated disease, *e.g.*, family history, genetic markers, early symptoms, and the like.

Immune system-mediated diseases that may be treated by the subject methods include, but are not limited to, *e.g.*, rheumatoid arthritis, osteoarthritis, reactive arthritis, autoimmune hearing disease, cartilage inflammation due to bacterial or viral infections (e.g. Lyme's disease), parasitic disease, bursitis, corneal diseases, and ankylosing spondylitis (fusion of the spine). The subject methods of the invention comprise the step of administering an effective amount of a composition of the invention, e.g., collagens, collagen derivatives. Preferred compositions for use in treating specific immune system-mediated diseases are fusion proteins of type IX collagen, preferably chimeras of a type IX collagen subunit with type II collagen and/or type XI collagen, and derivatives and subunits thereof, as well as as described in the preceding sections. In a preferred embodiment of the subject methods, the compositions administered to the subject comprise variably glycosylated collagens. The compositions administered in the subject methods are administered so that the active components, *i.e.*, collagens and/or collagen derivatives, contact the lymphoid tissue of the gut, *e.g.*, Peyer's patches or other similar sites, so that immune tolerance is induced. Such administration may be effected, by among many possible methods, through the use of formulations comprising the subjected compositions that are designed for oral

administration, *i.e.*, the active components are not destroyed or inactivated in the mouth, stomach, or other portions of the digestive system prior to contacting the appropriate gut lymphoid tissue. The treatment methods of the invention may also comprise the steps of administering additional pharmaceutical compounds for the treatment of immune system-mediated diseases, such as anti-inflammatory agents and the like.

The dosage at which the subject compositions are administered may vary within a wide range and will depend on various factors such as for example the severity of the inflammation, the age of the patient, etc., and may have to be individually adjusted. As a possible range for the amount of collagen(s) and/or collagen(s) derivatives which may be administered per day may be in the range of from about 0.001 mg to about 200 mg. Preferably, the amount of collagen and/or collagen derivatives administered is low, thereby favoring the induction of immune tolerance by suppression rather than clonal anergy. The pharmaceutical compositions containing the collagen(s) and/or collagen(s) derivatives may suitably be formulated so that they provide doses within these ranges either as single dosage units or as multiple dosage units.

The optimal dosage of tolerance inducing compositions for use in the methods of the invention will vary in accordance with a number of factors. The terms "dosage" and "dose" as used herein, unless indicated otherwise, may refer not only to a single administration of a composition but may be used to refer to the total amount of a given pharmaceutical composition administered over a selected period of time and involving multiple individual administrations. Factors affecting the optimal dosage include the choice of collagen molecule or molecules (and/or collagen derivatives) administered to the patient, the specific mucosa binding molecules selected, the age of the patient, the severity of the disease, other diseases that may be present in the patient, inert components in the formulation, adjuvants, and the like. There may be considerable variation in the range of dosages that are effective in treating a given immune disorder. Different dosages of the same pharmaceutical composition may produce the desired tolerance effect by different mechanisms. Although the operation of the invention is not dependent upon a particular theory of operation, the person of ordinary skill in the art will better understand the invention and provide additional embodiments by appreciating

that there are believed to be two primary mechanisms by which oral tolerance is mediated. Oral tolerance may be mediated by active cellular suppression in which regulatory T cells that suppress the activation and proliferation of lymphocytes specific for tolerized antigen. Another mechanism of oral tolerance induction is clonal anergy in which T lymphocytes having a suitable receptor are rendered unresponsive. Generally active suppression tolerance is favored by "low" doses of a tolerizing antigen and clonal anergy is favored by comparatively "high" doses of the same tolerizing antigen. A review of the principles and techniques for oral tolerance induction can be found in Weiner *et al.*, *Annual Review of Immunology*, pp. 809-835, Annual Reviews (1994).

The subject compositions may be formulated as pharmaceutical compositions so as to be adapted for certain types of administration to mucosal surfaces, e.g., oral, topical, and inhalation. The preferred form of formulation for oral administration in a form where the collagen and/or collagen derivatives in the composition come into contact with intestinal lymphoid tissue, *e.g.*, Peyer's patches. Compositions of the invention may be administered topically, orally, intranasally, by injection or by inhalation in the form of a pharmaceutical compositions comprising a collagen(s) and/or collagen(s) derivatives in the form of the original compound or optionally in the form of a pharmaceutically acceptable salt thereof, in association with a pharmaceutically acceptable carrier which may be a solid, semi-solid or liquid diluent or an ingestible capsule, and such preparations comprise a further aspect of the invention. The collagen(s) and/or collagen(s) derivatives and mucosa binding collagen conjugates may also be used with carrier material. As examples of pharmaceutical preparations may be mentioned tablets, drops such as nasal drops, preparations for topical application such as ointments, jellies, creams and suspensions, aerosols for inhalation, nasal spray, liposomes, etc. Usually the collagen(s) and/or collagen(s) derivatives will comprise between 0.05 and 99%, or between 0.1 and 99% by weight of the preparation, for example between 0.5 and 20% for preparations intended for injection and between 0.1 and 50% for preparations intended for oral administration.

To produce pharmaceutical preparations in this form of dosage units for oral application containing a compound of the invention the active ingredient may be mixed with a solid, pulverulent carrier, for example lactose, saccharose,

sorbitol, mannitol, a starch such as potato starch, corn starch, amylopectin, laminaria powder or citrus pulp powder, a cellulose derivative or gelatine and also may include lubricants such as magnesium or calcium stearate or a Carbowax™ or other polyethylene glycol waxes and are compressed to form tablets or cores for dragees. If dragees are required, the cores may be coated, for example, with concentrated sugar solutions which may contain gum arabic, talc and/or titanium dioxide, or alternatively with a film forming agent dissolved in easily volatile organic solvents or mixtures of organic solvents. Dyestuffs can be added to these coatings, for example, to distinguish between different contents of active substance. For the preparation of soft gelatine capsules consisting of gelatine and, for example, glycerol as a plasticizer, or similar closed capsules, the active substance may be admixed with a Carbowax™ or a suitable oil as e.g. sesame oil, olive oil, or arachis oil. Hard gelatine capsules may contain granulates of the active substance with solid, pulverulent carriers such as lactose, saccharose, sorbitol, mannitol, starches (for example) potato starch, corn starch or amylopectin), cellulose derivatives or gelatine, and may also include magnesium stearate or stearic acid as lubricants.

The compositions of the invention may also be formulated so as to provide a sustained release. By using several layers of the active drug, separated by slowly dissolving coatings sustained release tablets may be obtained. Another way of preparing sustained release tablets is to divide the dose of the active drug into granules with coatings of different thicknesses and compress the granules into tablets together with the carrier substance. The collagen(s) and/or collagen(s) derivatives and mucosa binding collagen conjugates may also be incorporated in slowly dissolving tablets made, for instance, of fat and wax substances or evenly distributed in a tablet of an insoluble substance such as a physiologically inert plastic substance.

In order to obtain dosage units of oral preparations -- tablets, capsules, etc. -- which are designed so as to prevent release of and possible decomposition of the active substance in the gastric juice, the tablets, dragees etc. may be enteric-coated, that is provided with a layer of gastric juice-resistant enteric film or coating having such properties that it is not dissolved at the acidic pH in the gastric juice. Thus, the active substance will not be released until the preparation reaches the

intestines. As example of such known enteric coatings may be mentioned cellulose acetate phthalate, hydroxypropyl-methylcellulose phthalates such as those sold under the trade names HP 55 and HP 50, and Edragit®L and Eudragit®S.

Liquid preparations for oral application may be in the form of elixirs, syrups or suspensions, for example solutions containing from about 0.1% to 20% by weight of active substance, sugar and a mixture or ethanol, water glycerol, propylene glycol and optionally aroma, saccharine and/or carboxymethylcellulose as a dispersing agent.

VI. EXAMPLES

The invention will be further understood by reference to the following examples, which are intended to be purely exemplary of the invention.

A. EXAMPLE 1: Expression Of Recombinant α3(IX) Collagen Subunit In *Pichia Pastoris*

PCR primers for the amplification of the α3(IX) collagen cDNA coding sequence from the plasmid p545 and the cDNA library clone RB410 are prepared. The primers are designed such that they introduce an Eco RI site at the 5' and the 3' termini of the α3(IX) collagen coding sequence, and a unique restriction site is used to join the two halves of the coding sequence found in these two clones.

A Primer 1 and a Primer 2 are used to amplify the mature amino-terminal coding sequence for α3(IX) collagen from plasmid p545 using standard PCR conditions as described in Ausubel *et al.*, *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley Interscience, N.Y. (1990). A Primer 3 and Primer 4 are used to amplify the remaining cDNA coding sequence, including the stop codon, from the cDNA clone RB410 as described above. The resulting PCR product is digested with the chosen unique restriction endonuclease and with EcoR I.

Commercially available expression vector pPIC9 (Invitrogen, San Diego, CA), which directs secreted expression in *Pichia pastoris* is digested with restriction endonuclease EcoR I, followed by calf intestinal phosphatase (Pharmacia), and then heat denaturation at 70°C for 5 minutes. The digested PCR

products and the pPIC9 vector are gel purified as described in Example 3 and a three-way ligation is performed. After transformation into competent *Escherichica coli*, correctly ligated plasmids are identified by restriction analysis and confirmed by sequencing using the commercially available *Pichia* sequencing primers (Invitrogen, San Diego, CA).

The α3(IX) *Pichia* expression vector is linearized and used to transform spheroblasts of a *his4 Pichia pastoris* strain which also expresses prolyl-4-hydroxylase. Transformants are identified on histidine deficient media and are confirmed by assaying for the loss of the AOX1 gene by slow growth on methanol media. Expression of the α3(IX) gene is induced by growing cells on methanol as the sole carbon source. α3(IX) collagen subunit protein is secreted into the growth medium and subsequently purified using standard centrifugation, filtration, and chromatographic techniques.

B. EXAMPLE 2: Expression Of Trimeric Human Type IX Collagen In *Pichia Pastoris*

In a similar manner, the *Pichia pastoris* strain which produces α3(IX) collagen subunit is engineered to coexpress the α1(IX) and α2(IX) collagen subunits in the same cell.

C. EXAMPLE 3: Expression Of Trimeric Human Type IX Collagen In *Spodoptera Frugiperda* Sf9 Insect Cells

Three recombinant viruses were generated by cotransfection of recombinant α1(IX), α2(IX), and α3(IX) constructs and a modified *Autographa californica* nuclear polyhedrosis virus DNA into *Spodoptera frugiperda* Sf9 insect cells using the Baculogold transfection kit (Pharmingen). The sequences used to construct the three a(IX) chains are disclosed in van der Rest and Mayne, *Structure and Function of Collagen Types* (Mayne, R. and Burgeson, R. eds.) Academic Press, Orlando, FL, pp. 185-221 (1987). The resultant viral pools were collected, amplified, and plaque- purified as described by Gruenwald, S. and Heitz, *J., Baculovirus Expression System: Procedures & Methods Manual*, Pharmingen, San Diego, CA (1993).

Spodoptera frugiperda Sf9 insect cells were cultured in TNH-FH medium supplemented with 10% fetal bovine serum (BioClear) as monolayers at 27°C. Approximately, 5 x 10^6 insect cells were infected with recombinant human α1(IX), α2(IX), and α3(IX) constructs and with the αβ virus for human prolyl 4-hydroxylase (manuscript in preparation). Viruses for the type IX collagen α-chains were used in 2 to 3 fold excess over the prolyl 4-hydroxylase virus. Ascorbate of 80 μg/ml was added to the culture medium daily. The culture media was removed 72 hours after infection, and the cell layer was washed once with a solution of 0.15 NaCl and 0.02 M phosphate at pH of 7.4. The cells were harvested by scraping them in 1.4 ml of ice cold solution of 0.5 acetic acid, 0.75 M NaCl, 10mM EDTA, and 1mM PMSF at pH of 2.5. The cells were then homogenized, and centrifuged at 15000 x g for 20 minutes. Supernatant was precipitated with NaCl at the final concentration of 1.2 M by mixing the sample for 12 hours at 4°C. The precipitate was centrifuged at 15000 x g for 20 minutes at 4°C. The resulting pellet was dissolved in 500 μl of cold 50mM acetic acid for three hours at 4°C. A sample of 15 μl was analyzed by either non-reducing or reducing SDS-PAGE followed by staining with Coomassie Brilliant Blue. The material was also digested with pepsin for 4 hours at 22°C, and the thermal stability of the pepsin-resistant recombinant type IX collagen was measured by rapid digestion with a mixture of trypsin and chymotrypsin as described in Buckner *et al, Anal. Biochem* 110:360-368 (1981). The resulting material was analyzed on reducing SDS-PAGE followed by Western Blotting with an antibody to a triple-helical collagen.

The results indicated expression of human type IX collagen as a heterotrimer of about 300 kDA (Figure 1) consisting of equal amounts of α1(IX), α2(IX), and α3(IX) chains as indicated in Figure 2. The thermal stability of the recombinant human type IX collagen was analyzed after a brief protease digestion. The thermal stability of the recombinant human type IX collagen was over 40°C.

D. EXAMPLE 4: Cloning And Expression Of A Chimeric Type II/IX/XI Collagen Molecule

The α3(IX) *Pichia* expression vector described above is modified so as to direct expression of a chimeric type II/type IX/type XI molecule.

Specifically, the vector is cleaved either 5' or 3' to the α3(IX) collagen coding sequence and coding sequence for type II collagen is inserted in-frame. Additionally, the vector may be again cleaved, either 5', 3', or between the type II and type IX coding sequence, and sequence encoding type XI collagen is also inserted in the correct reading frame so as to express a chimeric type II/IX/XI collagen molecule. Transformants of competent *E. coli* are screened for plasmids with the desired orientation by restriction digestion and confirmed by sequencing as described above.

Various modifications of the invention, in addition to those shown and described herein, will become apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall within the scope of the appended claims. It is also to be understood that all base pair sizes given for nucleotides are approximate and are used for purposes of description.

All references cited herein are hereby incorporated by reference in their entirety.

WHAT IS CLAIMED IS:

1. A fusion protein comprising a human type IX collagen linked to heterologous peptide sequence.

2. The fusion protein of Claim 1 wherein the heterologous peptide sequence comprises a type II collagen.

3. The fusion protein of Claim 1 wherein the heterologous peptide sequence comprises a type XI collagen.

4. The fusion protein of Claim 1 wherein the heterologous peptide sequence comprises type II and type XI collagen.

5. A method for producing recombinant human fusion protein comprising:

(a) culturing a host cell transformed with the recombinant DNA expression vector which expresses said fusion protein; and

(b) recovering the fusion protein from the cell culture.

6. A protein comprising human recombinant type IX collagen.

7. A method for producing recombinant human type IX collagen comprising:

(a) culturing a host cell transformed with the recombinant DNA expression vector which expresses said type IX collagen; and

(b) recovering the type IX collagen from the cell culture.

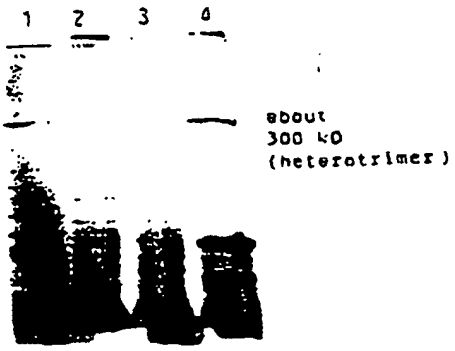

FIGURE 1

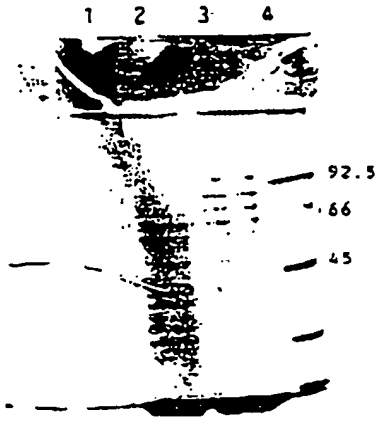

FIGURE 2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/18149

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :A61K 38/17; C07K 1/00, 14/00; C12P 21/00
US CL :530/356; 435/69.1, 69.7, 70.1, 71.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/356; 435/69.1, 69.7, 70.1, 71.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Dialog, APS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ADAMS et al. Collagen Gene Expression. American Journal of Respiratory Cell and Molecular Biology. 1989, Vol.1, pages 161-168, especially pages 161-166.	1-7
Y	MAYNE et al. New Members of the Collagen Superfamily. Current Opinion in Cell Biology. 1993, Vol.5, pages 883-890, especially pages 883-887.	1-7
Y	VAN DER REST et al. Collagens: Diversity at the Molecular and Supramolecular Levels. Current Opinion in Structural Biology. 1993, Vol.3, pages 430-436, especially pages 430-434.	1-7

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 09 JANUARY 1997	Date of mailing of the international search report 30 JAN 1997
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer Deborah Frelse ENRIQUE D. LONGTON Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/18349

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 93-07889 A1 (THOMAS JEFFERSON UNIVERSITY) 29 April 1993 (29/04/93), see entire document, especially pages 1-23.	1-7
Y	CALAMIA et al. Lac Permease of Escherichia coli: Topology and Sequence Elements Promoting Membrane Insertion. Proc. Natl. Acad. Sci. USA. July 1990, Vol.87, pages 4937-4941, especially pages 4937-4940.	1-5